DOCUMENT RESUME

ED 261 087 TM 850 511

AUTHOR Schafer, William D; Johnson, Charles E.
TITLE Microcomputer Support in Basic Statistics Instruction.
PUB DATE 85
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
AVAILABLE FROM William Shafer and Charles Johnson, Department of Measurement, Statistics, and Evaluation, College of Education, Univ. of Maryland, College Park, MD 20742 (Statistics Disk copies of software packages, $10.00).
PUB TYPE Speeches/Conference Papers (150) -- Computer Programs (101) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adaptive Testing; *Computer Assisted Instruction; Computer Assisted Testing; *Courseware; Higher Education; *Microcomputers; *Statistics; Student Evaluation
IDENTIFIERS Apple II Plus

ABSTRACT

This paper presents examples of effective uses of microcomputers to support basic statistics instruction. All programs are written in Applesoft BASIC for Apple II Plus microcomputers and compatible equipment. They have been field tested in statistics courses at the University of Maryland. Microcomputers can be used with color monitors for classroom demonstrations in conjunction with a lecture format for data presentation and to illustrate scatter diagrams, normal areas, binomial distribution and sampling from a beta population.A menu driven data handling and processing software program is used by students outside of class to provide practice in making data analysis decisions as well as an understanding of what users need to supply for storing, editing and processing data by computer. Computer-administered individualized diagnostic review testing provides performance feedback to students and performance data to instructors on optional quizzes as practice for required pencil and paper course quizzes and exams. Appendices contain listings for the scatter design, normal areas, and binomial programs, plus course assignments using the statistics package. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED261087

MICROCOMPUTER SUPPORT IN BASIC STATISTICS INSTRUCTION

William D. Schafer and Charles E. Johnson
University of Maryland

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Schafer, William D.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☑ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 850 511

BEST COPY AVAILABLE

Microcomputer Support in Basic Statistics Instruction

William D. Schafer and Charles E. Johnson
University of Maryland

The purpose of this paper is to present and discuss examples of effective uses of microcomputers to support basic statistics instruction. The paper is divided into three sections: (1) classroom demonstrations which illustrate concepts discussed by an instructor and are intended to be incorporated within the lecture format; (2) an educationally oriented data handling and processing package of programs to be used by students outside of class in order to provide practice in making decisions pertinent to analysis of data as well as an understanding of what users need to supply in order to accomplish storage, editing, and processing of data by computer; and (3) diagnostic review of concepts in statistics through individualized presentation of items in a quiz format with directed feedback depending on responses. All programs used in the illustrations are written in Applesoft BASIC and have been field-tested. Each of the three areas is amplified below.

Instructors who would like to use the statistics package may obtain copies on disk at cost, which may be copied freely (see page 6). Listings of several of the demonstration programs are included in appendices and other information is available from the authors as noted in the paper.

(1) Classroom Demonstrations

Several examples are described. Each is intended to be used in conjunction with a lecture presentation for purposes of illustrating concepts being discussed; none is a "stand-alone" instructional program. These examples do not cover all concepts in a basic statistics course but do illustrate a variety of ways in which a microcomputer, capitalizing on its graphics capabilities, can be used in class by an instructor to support a lecture presentation. Discussion of how each has been used is included. The equipment we have been using is an Apple II+ microcomputer connected to a 25-inch color monitor; our class sizes have been about thirty.

(a) Data Presentation

This set of programs can be used prior to class to enter and save data sets, such as from the textbook, into files for in-class use. It also allows random samples to be developed from uniform or normal populations. In use, the programs give the number of scores and the highest and lowest observed values. Grouping parameters are then requested. The computer then presents the frequency distribution, histogram, frequency polygon, and cumulative distribution graph corresponding to the grouping requested. The grouping parameters may then be changed to observe their impact on these forms of presentation for the given data set.

Paper presented at the 1985 convention of the American Educational Research Association, Chicago, Illinois. Partial support from the University of Maryland Computer Science Center is gratefully acknowledged.

We use these programs as we introduce data organization and display. The data used in a "blackboard" presentation are entered and processed so that students can validate, for themselves, the accuracy of the programs. We have also entered several data sets from the text so that students may turn to the appropriate page to see the raw data as they are being organized and presented on the screen. Then, different grouping parameters are requested from the class and their effects observed. The programs also give descriptive statistics and find percentiles from the data as requested by the user. A listing of the programs may be obtained from the authors.

(b) Scatter Diagram

This program presents a scatter diagram for a sample of user-defined size from a bivariate normal (approximate) with user-defined correlation. It also plots the population regression line on the diagram.

We use this program to illuatrate the natures of scatter diagrams for varying correlation coefficients. The students can supply coefficients and then be asked to predict the form of the diagram. Since the program executes rapidly, many examples may be developed in a short time. A listing of this program may be found in Appendix A.

(c) Normal Areas

This program was written by Randall E. Krack while a student in one of our microcomputer programing courses. Input is the mean and standard deviation of a normal and the lower and upper limits of the area desired, either of which may be "infinity." If infinity is given, the program substitutes + or - 4 as the standard score. It plots the portion of the normal so identified with vertical lines at integer standard score values superimposed and gives the proportion of area subsumed.

We use this program first when introducing finding normal areas. Since the program highlights graphically the area to be found, it becomes a valuable visual aid in the presentation. Students may check their values for the areas found using the standard normal table during their work with practice examples against the output of the program.

We re-use this program when we cover power for z tests on means. The "null" sampling distribution's region of rejection may be graphed to show that the proportion of that region is, indeed, alpha. Then the region of rejection may be graphed to show the power for any "alternate" normal sampling distribution. The graph of the portion of the normal in the region of rejection provides a visual with strong impact to help illustrate the lecturer's points. A listing of this program may be found in Appendix B.

(d) Binomial Distribution

Binomial probabilities are given and graphed with a histogram for user-specified probability (pi) and sample size (n). The graphs may be used to illustrate the effects of these parameters on such characteristics as power of the binomial test and to illustrate their effects on the binomial shape. Since the algorithm for computing binomial probabilities is quite fast, even sample sizes above 100 may be shown quickly, allowing demonstration that the binomial shape approximates normality as n increases. The mean and standard deviation of

the binomial are also displayed on the screen.

We introduce hypothesis testing concepts using the binomial test as the vehicle, so this program is used often. With a constant n, the horizontal scaling remains the same for any choice of pi, so the portions of the sampling distribution in regions of rejection and of retention may be demonstrated for different values of pi graphically. A listing of this program may be found in Appendix C.

(e) Sampling from a Beta Population

Following input of the two beta-distribution parameters, the beta is graphed; the parameters may then be changed if desired. The scale of the domain is placed on the interval (0-100). The program then accumulates a histogram on a sample-by-sample basis of the means of N random samples of size n each, where N and n are user-specified. The theoretical and observed means and standard deviations of the resulting sample from the sampling distribution of the mean are given.

This program may be demonstrated in class, but is too slow for repeated uses. However, it may be used individually by students to see the characteristics of sampling distributions of the mean from a variety of population distribution characteristics. A listing of this program may be obtained from the authors.

(2) A Menu-Driven Statistics Package

The use of a menu-driven microcomputer package for processing data allows students to focus attention on the decisions involved in selecting an analysis and in interpreting the results rather than on either the computations or the structure of a package's language. The microcomputer package we developed for introductory statistics students has several features which we feel are desirable in an instructional setting.

(a) Description of the Package

The package is entirely menu-driven, using two disks, one for the package of programs and the other for storage of data files. The package disk is write-protected and copies are checked out for student use in the College of Education microcomputer laboratory. Each student is expected to purchase a data disk; instruction is given on how to initialize these. We have found that a twenty-minute classroom presentation is sufficient for students to be able to initialize their disks and use the package.

Two forms of data files are supported: "standard," which contain variable names and, if used, missing value codes (all statistical routines in the package are run from standard files), and "non-standard," which contain only lines of data. Either form may be used for data entry (the standard form is prompted), or data files may be transferred to the data disk in non-standard format from another computer using telecommunications software (not part of the package). A program in the package accomplishes the conversion to standard format by requesting variable names, column-position ranges, and, if used, missing-value codes. Thus, an instructor may electronically transfer and use data files existing elsewhere if the telecommunication facilities are available. Files

larger than may be stored in the microcomputer's memory are segmented automatically into scratch files on the data disk and then merged, so the file size limitation depends only on disk space availability.

The package allows the user to:

create a standard file

create a non-standard file

see the directory of files on a data disk

copy a file, either on the same disk or on a second disk (this feature allows students to copy one or more common files from the program disk, itself, onto their data disks if the instructor desires)

add cases to a file

add a variable to a file

convert a file from non-standard format to standard format

delete a file

edit either type of file

run a statistical procedure.

In editing a standard file, the user may:

change a variable's name

add, delete, or change missing value codes

locate and replace a value of a variable for a case

sort the file on a variable

list the file, with the option of sending the listing to a printer

delete cases

delete variables.

A separate menu presents these options once the standard file editor is selected from the initial menu.

The statistical procedures available are:

frequency distribution

descriptive statistics

one-way chi-square

two-way chi-square

regression analysis

intercorrelations

one-group t-test

two-group t-test

analysis of variance

scatter diagram.

Categorical variables may be described by codes or by ranges of values with names associated with the categories in either case. The user chooses the analysis from a menu and is then asked to describe the variables to be used as appropriate for that analysis.

Both one-disk-drive and two-disk-drive versions of the package are available for Apple II+ and compatible equipment. An attempt has been made to anticipate all possible errors a student may make and provide for brief comments leading to correct re-entry of information. In practice, students very rarely exit the system and go to command level through error, and all instances which the computer lab personnel have brought to our attention have been corrected.

Since many microcomputer laboratories have printers attached only to some equipment, as does ours, the package allows the student to re-run the last analysis set up on the data disk automatically as an option. Thus, once the analysis is satisfactory, the two disks can be taken to a microcomputer equipped with a printer and the analysis repeated for printing without setting it up again. All output is in text format, only, so that no capability is necessary for special graphics output. As used, students run homework assignments using the package and hand in the output, annotated as required in the assignments. The text of the assignments as handed out to the students may be found in Appendix D.

(b) Use of the Package

There are three data sets used for the assignments: (1) a set entered by the instructor and copied onto their disks by the students containing measures of educational variables on each of the 50 states (this data set is included on all copies of the package), (2) a set entered by each student from a table in the textbook (Elifson, Runyon, and Haber, 1982), and (3) a set entered by students based on the results of surveys they design and administer to the class. We have twelve copies of the one-drive-version and two copies of the two-drive-version available for checkout; this has been sufficient for approximately 200 students in six sections of the course in our heaviest semester. The package is now in its third semester of use and is performing well with an even greater demand.

(c) Student Reactions

At the end of the semester, students were asked to evaluate the package. The results of the 139 responses received from the seven sections, taught by various instructors, are given below in percentages. The response options were

Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), and Strongly Disagree (SD).

Item	SA	A	N	D	SD
The computer assignments were appropriate for this course.	32	52	9	5	2
The computer package is relatively easy to use.	43	50	6	0	2
Students became more likely to use a computer for statistics, if available.	32	42	20	5	1
Students became more knowledgeable about computer data analysis.	21	48	21	7	2

(d) Availability of the Package

Instructors interested in using the package may obtain a copy on disk of both versions for a materials and postage charge of $10. Multiple copies may be made with standard duplication software (i.e., COPYA). Send your request to the authors at Department of Measurement, Statistics, and Evaluation, College of Education, University of Maryland, College Park MD 20742. Make checks payable to University of Maryland.

(3) Individualized Diagnostic Review Testing

Computer-administered tests can provide individualized performance feedback to students and can provide instructors with various types of performance data. This paper describes our implementation of such a computer-testing system.

A set of programs coordinates the use of individual, computer-administered quizzes to students. The quizzes are made available on an optional basis to provide performance feedback and practice for required paper-and-pencil course quizzes and exams. Multiple copies of quiz disks are made available in the computer lab for students to check out and use in the lab at their convenience. Several features of the system will be presented, followed by some evaluative comments.

(a) Programs

A single menu-driven utility disk contains programs for instructors' use in item development, quiz administration, and data management. Progams on the utility disk are used to develop (1) item file disks for storing or banking items, (2) quiz disks for computer quiz administration, and (3) data disks for storing quiz results. Selected item files are copied to quiz disks so that quizzes may be administered using a single disk drive.

(b) Item Development and Banking

Multiple-choice and completion items are input (may be mixed in the same file) along with single or multiple keys. Statistical symbols can be used by

incorporating a commercial graphics routine allowing alternate character sets. Items can be edited on input or on recall later from a disk file. They can be copied within and between files to facilitate building new items from old (e.g., when the same tables are used for different items). Item files are in random-access format with a maximum size of 500 bytes. A disk can hold about 250 items allocated among user-defined files.

(c) Specification of Quizzes

Several options are used to specify quiz content and the process of quiz administration. These include: (1) topics to be presented, (2) whether or not students may freely choose from a topic menu, (3) the number of items to be administered per topic (default is all), (4) whether items are to be selected by the inbstructor or randomly selected from a specified item file, (5) whether randomly selected items are to be across or within students, and (6) whether students have the option to choose their own stopping points. A specification file is written on a quiz disk and read by the quiz program.

(d) Quiz Administration

Quizzes are automatically administered when a quiz disk boots. Students are asked to enter ID information and test instructions are presented. Except when otherwise specified, students are presented a menu of quiz topics from which to choose. Items are then presented (fixed or randomly selected) according to instructor specifications using an answer-until-correct procedure. Following incorrect responses, students are asked to try again. If after three tries the answer is still incorrect, the item is presented along with the correct answer. When a student chooses to quit or completes the specified number of items, ID information and response data are written to a data file on the quiz disk and a performance summary is presented along with the opportunity to review missed items. Students may then select another topic from the topic menu.

(e) Managing Data

Utility programs are used to observe and print response data and to copy data from individual test disks to a summary data disk for later instructor review and analysis. A student ID file can be produced and used by the quiz program in order to screen students taking quizzes or to assign students to experimental variations of quiz conditions.

(f) Evaluation

Computer quizzes have been available over several semesters to students in two graduate-level introductory research and statistics courses. They have also been used in two experimental studies of types of item presentation. There have been very few disk or mechanical problems in these applications, although some students have complained of distractions in the heavily used computer lab.

In a survey of eighty-five (86%) of 99 graduate students in an introductory statistics course, the mean number of administrations was 4; with a total of 343 separate administrations. No quiz disk malfunctions were reported and only one case of machine malfunction. The value of the practice quizzes was reported as: very high (26;34%), high (25;32%), neutral (9;12%), low (5;6%), very low (12;16%) When asked if they would have made use of additional quizzes had they been available responses were: very likely (37;47%), likely (15;19%), undecided

(5,6%), unlikely (8;10%), very unlikely (13;17%).

A major limitation has been the development of item files of adequate size and quality. Efforts are underway now to improve the value of quiz feedback (e.g., by enabling students to obtain a printed summary of areas of difficulty) and considering the incorporation of a mastery decision model.

(g) Availability

Persons who wish to receive further information about these programs may write to the authors.

Appendices

Appendix A: Listing of the Scatter Diagram Program

```
1  REM  SCATTERGRAM
2  TEXT : HOME
3  PRINT : PRINT : PRINT : PRINT : PRINT
10  PRINT  TAB( 8)"PROGRAM SCATTERGRAM"
12  PRINT : PRINT : PRINT
14  PRINT "THIS IS A DEMONSTRATION OF A BIVARIATE DISTRIBUTION": PRINT
16  PRINT "RANDOMLY SELECTED POINTS ARE PLOTTED FOR ANY LEVEL OF CORRELATION YOU
SELECT": PRINT
20  PRINT : PRINT : PRINT
30  INPUT "THE CORRELATION COEFFICIENT IS -- ";R
40  PRINT : PRINT : PRINT
50  INPUT "THE SAMPLE SIZE IS -- ";N
60  FOR I = 1 TO 500
65 PAUSE = I: NEXT I
70  HOME
90 R =  - R
100  HGR
110  COLOR= 2
112  HPLOT 0,0 TO 0,159 TO 279,159
118 B = R * .7
119 A = 80.5 - B * 150.5
1000  FOR I = 1 TO N
1005 RDAV = ( RND (1) +  RND (2) +  RND (3) +  RND (4) +  RND (5)) / 5
1010 X = RDAV * 200 + 50
1020 YP = B * X + A
1022 RDAV = ( RND (1) +  RND (2) +  RND (3) +  RND (4) +  RND (5)) / 5
1025 EE = RDAV - .5
1030 E = EE * (1 - R ^ 2) ^ .5 * 140
1040 Y = YP + E
1041  IF Y ( 0 GOTO 1005
1042  IF Y ) 159 GOTO 1005
1050  HPLOT X,Y
1060  NEXT I
1062  VTAB 22
1065  INPUT "PRESS RETURN TO OBTAIN THE POPULATION REGRESSION LINE. ";GO$
1070 YL = B * 51 + A
1080 YR = B * 250 + A
```

```
1090  HPLOT 51,YL TO 250,YR
1092 R =  - R:B =  - B
1095  PRINT : PRINT
1096  PRINT "R = "R"   N= "N
1097  PRINT
1098  INPUT "TYPE RETURN TO CLEAR";GO$
1099  TEXT : HOME
```

Appendix B: Listing of the Normal Areas Program

```
90  TEXT : HOME
100  PRINT "*****************************"
110  PRINT "*  NORMAL GRAPH             *"
120  PRINT "* COMPUTES AREA UNDER      *"
130  PRINT "* CURVE AND GRAPHS         *"
140  PRINT "*****************************": PRINT : PRINT
150  PRINT "PROGRAMMED BY RANDALL E. KNACK"
160  PRINT : PRINT "     MARCH, 1984"
170  PRINT : PRINT
180  INPUT "PRESS RETURN TO CONTINUE";A$
190  HOME
200  PRINT "YOU MUST PROVIDE THE UPPER AND LOWER LIMITS OF X, THE MEAN OF X AND
THE  STANDARD DEVIATION OF X."
210  PRINT
230  PRINT
240  INPUT "INPUT MEAN OF X = ";MX
250  PRINT
260  INPUT "INPUT STANDARD DEVIATION OF X = ";SD
270  PRINT
275  PRINT "IF THE UPPER OR LOWER LIMIT OF X IS -INFINITY OR +INFINITY INPUT ONLY
THE WORD ": INVERSE : PRINT "INFINITY": NORMAL
276  PRINT
280  INPUT "INPUT LOWER LIMIT OF X = ";A$
290  IF A$ (  ) "INFINITY" THEN 310
300 ZL =  - 4: GOTO 320
310 ZL = ( VAL (A$) - MX) / SD
320  PRINT "Z (LOWER LIMIT) = ";ZL
330  PRINT : INPUT "UPPER LIMIT OF X = ";A$
340  IF A$ (  ) "INFINITY" THEN 360
350 ZU = 4: GOTO 370
360 ZU = ( VAL (A$) - MX) / SD
370  PRINT "Z (UPPER LIMIT) = ";ZU
380  PRINT : PRINT "IS YOUR INPUT CORRECT?"
390  PRINT "IF IT IS NOT, TYPE NO AND YOU CAN DO IT AGAIN."
400  PRINT "PLEASE USE YES OR NO"
410  INPUT "IS DATA CORRECT? ";A$
420  IF A$ = "NO" THEN  GOTO 190
490  HOME
500  REM  *****  FILL IN DISTRIBUTION FROM LL TO UL  *****
510  REM  RESCALE Z RANGE TO DISPLAY VALUES AND ROUND TO INTEGERS
520 RL =  INT (ZL * 28 + 140.5)
530 RU =  INT (ZU * 28 + 140.5)
545  HGR
550  HCOLOR= 1
```

```
560  FOR I = RL TO RU
561 A = (I - 140) ^ 2
562 B = A / 1568
563 C =  EXP ( - B)
570 HT = C * 70.1855923
580 HT =  INT (HT + .5)
584 HT = 135 - HT
590  HPLOT I,135 TO I,HT
600  NEXT I
700  REM  ****DRAW BASELINE AND DEVIATION SECTORS LAST *****
710  HCOLOR= 5
750  REM  RI=Z SCORE VALUES -4 TO +4
760  FOR I =  - 4 TO 4
770 RI = I * 28 + 140
780  HPLOT RI,10 TO RI,145
781  HPLOT RI + 1,10 TO RI + 1,145
782  HPLOT RI + 2,10 TO RI + 2,145
790  NEXT I
960  REM ****   CALCULATE AREA UNDER CURVE USING SIMPSON'S RULE    ******
970  REM    ***  REFERENCED IN TOM RUGG & PHIL FELDMAN'S TEXT...ADAPTED BY REK
***
980  REM   *  NORMAL CURVE FUNCTION = Y *
990 P = 0:N = 2: GOTO 1020
1000 Y = .3989423 * ( EXP ( - 1 * (X ^ 2 / 2)))
1010  RETURN
1020 DX = (ZU - ZL) / N:T = 0
1030 X = ZL: GOSUB 1000:T = T + Y
1040 X = ZU: GOSUB 1000:T = T + Y
1050 M = N / 2:Z = 0
1060  FOR J = 1 TO M
1070 X = ZL + DX * (2 * J - 1): GOSUB 1000:Z = Z + Y
1080  NEXT J:T = T + 4 * Z
1090 M = M - 1: IF M = 0 THEN 2030
2000 Z = 0: FOR J = 1 TO M
2010 X = ZL + DX * 2 * J: GOSUB 1000:Z = Z + Y
2020  NEXT J:T = T + 2 * Z
2030 A = DX * T / 3
2040 A =  INT (10000 * A + .5) / 10000
2050  IF A = P THEN 3000
2060 P = A
2070 N = N * 2
2080  GOTO 1020
3000  REM  PRINT
3005  VTAB 22
3010  PRINT "Z VALUES ARE GRAPHED"
3020  PRINT "PROPORTION OF TOTAL AREA= ";P
```

Appendix C: Listing of the Binomial Program

```
10  REM  BINOMIAL
45  REM  ******************
50  PRINT : INPUT "SAMPLE SIZE (NUMBER OF TRIALS): ";N
60  PRINT : INPUT "PROBABILITY OF A SUCCESS (PI): ";P
65  PRINT
```

```
70  DIM PX(N),CX(N)
75  PRINT "# SUCCESSES","PROB.","CUM. PR."
80 PC = 0
85 MX = 0
90 CX(0) = 1:CX(N) = 1
95 J =  INT (N / 2 + .005)
100  FOR Z = 1 TO J
105 CX(Z) = CX(Z - 1) * ((N - Z + 1) / Z)
110  NEXT Z
115  FOR Z = J + 1 TO N - 1
120 CX(Z) = CX(N - Z)
125  NEXT Z
140  FOR Z = 0 TO N
150 PX(Z) = CX(Z) * P ^ Z * (1 - P) ^ (N - Z)
158  IF PX(Z) ) MX THEN MX = PX(Z)
160 PC = PC + PX(Z)
170  PRINT Z, INT (10000 * PX(Z) + .5) / 10000, INT (10000 * PC + .5) / 10000
180  NEXT Z
182  INPUT "PRESS 'RETURN' FOR HISTOGRAM";A$
187  HOME
190  HGR
192  VTAB (23)
195  PRINT "N = ";N;"     PI = ";P
196  PRINT "MEAN = ";N * P;"   S.D. = "; SQR (N * P * (1 - P))
200  HCOLOR= 3
210  HPLOT 0,159 TO 279,159
215 WD =  INT (279 / (N + 1))
230 X =  - WD
240  FOR Z = 0 TO N
250 X = X + WD
260 Y = 159 - 159 / MX * PX(Z)
270  HPLOT X,159 TO X,Y TO X + WD,Y TO X + WD,159
310  NEXT Z
335  INPUT "PRESS 'RETURN' TO END";A$: TEXT
340  END
```

Appendix D: Course Assignments using the Statistics Package

Computer Assignments

The computer-processing aspects of the assignments are to be carried out using the Apple II+ computers in the Educational Technology Center in the Education Building. Several copies of the main package disk are available for room use. Always begin by booting a main package disk. A few of the computers in the Educational Technology Center are equipped with two disk drives. If you are using one of these, you may ask for a copy of the two-drive version of the main package disk; this will avoid the annoyance of disk swapping necessary with the one-drive version.

Most computer assignments use a data set titled EDUCATION which you are to copy from a main package disk to your data disk (each student should obtain a 5 1/4" floppy disk to do the assignments - the bookstore carries this item). The variables in this data set are:

No.	Name	Description
1	REGION	1=East 2=Midwest 3=West
2	ECONOMY	1=Natural Resources 2=Industry
3	GOVERNOR	1=Democrat 2=Republican
4	INCOME	Per-Capita Income of Residents of the State
5	PCTEDUC	Percent of State Budget for Education
6	GRADES	Mdn. Education Beyond 12th Grade (1/10 Years)
7	STUDENTS	Size of Largest State Univ. Campus
8	FEES	Fees and Tuition at that Campus
9	LIVING	Room and Board Costs at that Campus
10	STRATIO	Student-Faculty Ratio at that Campus

There are fifty cases in the file, one per state. Data are from the 1983 Reader's Digest Almanac, except for variable No. 6, which is from the Census Bureau's 1980 data. For two states (Maryland and Virginia), room and board costs were unavailable and are coded as missing in the data set.

Assignment specifications are:

1. Initialize your data disk following the instructions given in class. Then enter the data in Table 3.1 on page 52 of the text into a standard file (option CS from the package's main menu). Use two variables, the column in which the age appears (1 through 10) and the age, itself. Also, use the disk-to-disk copying routine (option DD from the package's main menu) to place a copy of the file EDUCATION on your data disk. There is nothing which must be turned in to the instructor for this assignment, but try to complete it by the due date as all other assignments except the team project use these two files.

2. Using the data from the file you entered, reproduce the grouped frequency distribution on page 55 of the text. You will need to use the RS option from the package's main menu and the name of the file you will process is the name you used for the text data set in assignment 1. You will need to define the categories by ranges, which should be chosen to correspond to the grouping intervals on page 55 of the text. Note that you have already entered the data; there is no need to enter data again! Also, obtain the distribution of the states by region in the EDUCATION data set. Here the categories may be defined by codes. Print both the frequency distribution and the histogram for each. By hand, draw a frequency polygon for the ages and a pie diagram for the regions.

3. Obtain descriptive statistics for both variables in the text data set and for the last seven variables in the EDUCATION data set. For each variable, by hand, compute its range and divide the result by its standard deviation. Circle the largest and smallest of these quotients.

4. Using the EDUCATION data set, obtain the scatter diagram with percent of state budget allocated to education on the horizontal axis and tuition and fees on the vertical axis. Also, find the intercorrelations among the last seven variables in the EDUCATION data set. Finally, produce the scatter diagram for the pair of variables which show the strongest association.

5. Obtain the regression equation to predict room and board expenses from tuition and fees. Tuition and fees at the University of Maryland are 1,185 dollars. Use the equation to estimate room and board expenses at UM. Obtain a figure for the current room and board expenses at UM (give your source) and

compare it with the calculated value. Comment on any reason(s) you might have to explain the discrepency.

6. The average age of the United States population (median) is 30.3 years. Use the data set from the text to test the hypothesis that the age of park visitors is different from the age of the national population. Also, test the hypothesis that the tuition and fees of the University of Maryland (1,185 dollars) is different from the average of the 50 states. For each test give the null and alternate hypotheses and a graph of the null distribution of the test statistic with the .05 region of rejection shaded and a dark point at the statistic's observed value. State whether or not the null is rejected.

7. Test the hypothesis that the student-faculty ratio differs according to the political party of the governor. Give the null and alternate hypotheses and a graph of the null distribution of the test statistic with the .05 region of rejection shaded and a dark point at the statistic's observed value. State whether or not the null is rejected.

8. Test the hypothesis that tuition and fees differs by region of the country. Also, test the hypothesis that age of park visitors differs according to the column in the original table. For each test give the null and alternate hypotheses and a graph of the null distribution of the test statistic with the .05 region of rejection shaded and a dark point at the statistic's observed value. State whether or not the null is rejected. Also, produce by hand the usual ANOVA summary table for each of these analyses using the information contained in the printout.

9. Test the hypothesis that the states in the regions of the country are equally numerous. Also, test the hypothesis that Democratic governors outnumber Republican by a ratio of three-to-one. For each test give the null hypothesis and a graph of the null distribution of the test statistic with the .05 region of rejection shaded and a dark point at the statistic's observed value. State whether or not the null is rejected.

10. Test the hypothesis that the governor's political party differs by the state's economic base. Also, grouping the GRADES variable into the intervals (1-4) (5) (6-8), test the hypothesis that the educational attainment of U. S. adults differs by region of the country. For each test give the null hypothesis and a graph of the null distribution of the test statistic with the .05 region of rejection shaded and a dark point at the statistic's observed value. State whether or not the null is rejected.

Project Assignment

Each project is to be completed by a group of four or five students. Each group is to collect as a minimum the following data from the class: two categorical (grouping) variables, one with at least three levels, and two continuous variables of interest to the group (class time is scheduled for this purpose - see the syllabus).

Your proposed procedures are to be turned in to the instructor for approval prior to the data collection (see the syllabus).

For each categorical variable describe its distribution with these techniques: (a) frequency distribution and (b) bar graph.

For each continuous variable produce: (a) frequency (or grouped frequency) distribution, (b) histogram, (c) frequency polygon, (d) cumulative frequency polygon, (e) mean, median, variance and standard deviation, and (f) a description of the nature (shape) of the distribution.

Analysis of the relationships between all six pairs of the variables should be carried out as follows:

(a) for continuous with continuous variables give the scatter diagram, the correlation coefficient along with its t test, and the regression equation corresponding to the scatter diagram;

(b) for categorical with continuous variables give the means and standard deviations of each group and a test for the significance of the differences between group means (t or ANOVA with HSD Tests, as appropriate);

(c) for categorical with categorical variables, give the two-way contingency table and the chi-square test for independence.

When presenting hypothesis tests, show null and alternate hypotheses, calculated statistic, its critical value, and a statement of the conclusions.

The final written project should include an introduction describing what variables you are studying, why you picked them and what results you expected to obtain, a description of how the data were collected and the characteristics of the subjects, a report of the results of all the analyses noted above, and some conclusions based on the results along with a discussion of why they might have turned out the way they did.

Calculations may be done most easily using computers. By entering the data into a file once you obtain it you may complete the project using the methods you have already learned for computer data analysis.

It is up to each team to divide the workload equitably. An oral report summarizing the results is to be presented to the class (see the syllabus).

Reference

Elifson, Kirk W., Runyon, Richard P., and Haber, Audry. Fundamentals of Social Statistics. Reading, Mass.: Addison-Wesley, 1982.